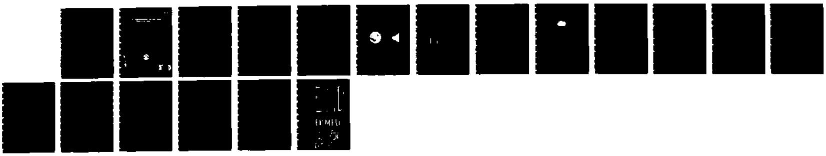
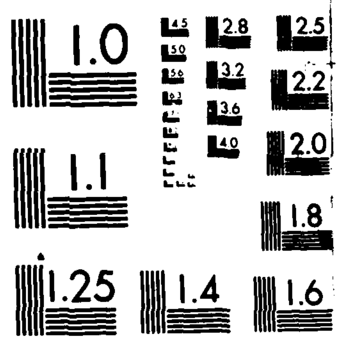
AD-A163 158 NARROWBAND INTEGRATED VOICE DATA SYSTEM BASED ON THE 2400-B/S LPC (LINEAR PREDICTIVE CODER)(U) NAVAL RESEARCH LAB WASHINGTON DC G S KANG 30 DEC 85 NRL-8942

1/1

UNCLASSIFIED F/G 17/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

NRL Report [illegible]

AD-A163 158

[illegible] on the 2400-b/s LPC

GEORGE S. KANG

Communication Systems Engineering Branch
Information Technology Division

December 30, 1985

FILE COPY

NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release; distribution unlimited.

DTIC
ELECTE
JAN 15 1986
S D
A

86 1 15 070

SECURITY CLASSIFICATION OF THIS PAGE

AD-A163 158

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Report 8942	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory — 6b. OFFICE SYMBOL (If applicable) 7526	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Space and Naval Warfare Systems Command — 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Washington, DC 20363-5100	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N		RR021-05-42	DN280-209

11. TITLE (Include Security Classification)

Narrowband Integrated Voice Data System Based on the 2400-b/s LPC

12. PERSONAL AUTHOR(S)
Kang, George S.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 10/84 TO 7/85	1985 December 30	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Narrowband LPC , Simultaneous voice/data transmission
			Integrated voice/data systems

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report presents a narrowband integrated voice/data systems which is achieved by modifying the Government-standard linear predictive coder (LPC) operating at 2400 bits per second (b/s) to transmit both speech and digital data simultaneously. The presence of the digital data in the speech bit-stream, however, does not cause operational incompatibility with other Government-standard LPCs which do not have this capability. Since the LPC will be widely deployed in tactical platforms to support command and control of the armed forces, the capability of transmitting simple graphics or typed messages along with continuous conversation would enhance the effectiveness of tactical voice communications. According to our experimentation, digital data up to approximately 75 b/s can be transmitted during continuous conversation without deteriorating the intelligibility of the LPC processed speech.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL George S. Kang	22b. TELEPHONE (Include Area Code) (202) 767-2157 — 22c. OFFICE SYMBOL 7526

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

CONTENTS

QUALITY INSPECTED 3

Accesion For
NTIS CRA&I
DTIC TAB
Unannounced
Justification
By
Distribution
Availability Codes
Dist | Avail and/or Special
A-1

NARROWBAND INTEGRATED VOICE/DATA SYSTEM BASED ON THE 2400 b/s LPC

INTRODUCTION

Until recently, telephones and telephone lines have been used exclusively for transmitting voice signals. In this computer age, however, telephone facilities are increasingly being used for transmitting digital data and graphics. The recently developed circuit switched digital capability (CSDC) by AT&T Laboratories is an example of new technology that will enable telephone companies and AT&T to support a host of digital services, including: bulk data transfer, teleconferencing, encrypted voice, and facsimile data. Even with new equipment, speech and digital data are not intended to be transmitted simultaneously.

The telephone was originally designed to transmit speech in analog form, and it is virtually impossible to transmit simultaneously both analog speech and digital data over a channel having a bandwidth barely large enough for the voice spectrum. On the other hand, the linear predictive coder (LPC) operating at 2400 bits per second (b/s) is a telephone which transmits voice in a digitized form. (Unless stated otherwise, the 2400-b/s LPC is referred to as the Government-standard 2400-b/s LPC defined by Federal Standard 1015 [1].) Originally, the 2400-b/s LPC was also designed to transmit only voice. Recently, however, a data transmission capability has been incorporated in certain 2400-b/s LPC systems. One example is the advanced narrowband digital voice terminal (ANDVT) which was developed under the technical direction of the Navy for triservice tactical use. ANDVT, however, was not designed to transmit voice and data simultaneously. Prior to establishing a communication link, the operator must choose either the voice or data mode.

Thus at present, there is no integrated voice/data system which is capable of transmitting voice and data simultaneously over a narrowband channel having approximately a 3 kHz bandwidth. The objective of the task described in this report was to develop such a system based on the 2400-b/s LPC. The most striking feature of our integrated voice/data system is that it interoperates with other 2400-b/s LPCs which do not have this capability.

We implemented a narrowband integrated voice/data system using a programmable 2400-b/s LPC at NRL. This system can transmit digital data at a rate up to approximately 75 b/s on the average during continuous conversation. Despite the presence of extraneous digital data in the bit-stream, speech intelligibility is unimpaired when evaluated by the diagnostic rhyme test (DRT). This report discusses the necessary software modifications to convert the 2400-b/s LPC to an integrated voice/data system, and it also shows copies of text and graphics received in real time while continuous radio news was processed by the LPC.

This report is a result of our effort to make the 2400-b/s LPC more acceptable to general users as a means of voice communications. Previously, we have improved the quality of LPC-processed speech through modifications to the voice processing algorithm [2,3]. We also introduced the delayed sidetone in the 2400-b/s LPC in an attempt to make the speaker talk at a slower rate so that LPC-processed speech could be better understood by the listener [4]. Also, we modified the bit-stream of the 2400-b/s LPC to provide a simultaneous voice and data transmission capability. We anticipate that the transmission of a few written headlines and simple graphics will enhance the effectiveness of voice communications, particularly in a military environment.

Manuscript approved August 30, 1985.

BACKGROUND DISCUSSIONS

Voice communication plays a vital role in command and control of the armed forces (Fig. 1). The 2400-b/s LPC will be deployed in the air, sea, and mobile ground tactical platforms to provide secure voice communications over narrowband channels having approximately a 3 kHz bandwidth, such as telephone lines and high frequency (HF) channels and satellite links. In a few years, the 2400-b/s LPC will also be used extensively by civilian agencies for secure office-to-office communications. Because narrowband channels are more readily accessible to general users and less expensive to lease than wideband channels, a majority of secure voice communications is expected to use the 2400-b/s LPC.

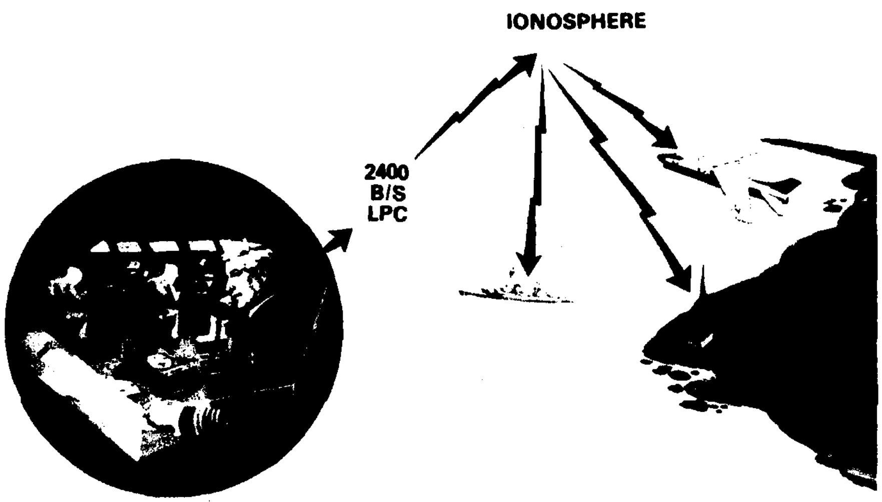


Fig. 1 — The 2400-b/s LPC operating over HF links

An advantage of two-way speech communication is its speediness and immediacy. As a means of communications, however, speech has some limitations which are not associated with writing:

- Casual speech is composed during the course of speaking. Hence, it often has a loose grammatical structure. Writing is usually better structured because it is composed under premeditation.

- Speech, a stream of sound, must be received by the listener as it comes. Thus deciphering speech in real time requires a constant mental concentration by the listener. Writing may be read again as needed because it is not bound to the time continuum.

- Speech communications are adversely affected by the presence of severe background acoustic noise at the speaker's site and/or receiver's site, whereas written messages are not degraded by acoustic noise interference.

As previously mentioned, speech and writing can complement each other. This is the reason why we use viewgraphs (written highlights or illustrations) during oral presentations. In the absence of visual aids, even hand-waving can improve communication. It is interesting to recall that when the telephone was first introduced at the turn of the century, many people had considerable difficulty communicating over the telephone because they could not see hand-wavings or facial expressions. This exemplifies the fact that a small amount of additional information related to speech could make voice communication more effective.

At present, however, none of the analog or digital telephones operating over narrowband channels is capable of transmitting written words or simple illustrations. Such a capability would be highly desirable for the 2400-b/s LPC because it is a difficult device to talk over (Fig. 2). The task described in this report is motivated by such a need.

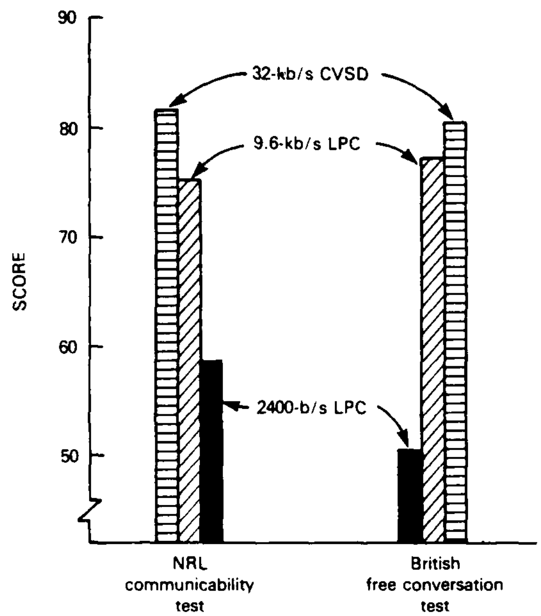


Fig. 2 — Two conversational test scores for various voice encoders, 32-kilobits per second (kb/s) continuously variable slope delta (CVSD) encoders, 9.6-kb/s residual-excited LPC developed by NRL, and the DoD-standard 2400-b/s LPC. This figure implies that the 2400-b/s LPC is not an easy device to talk over. The telephone scores in the lower 90s. In the NRL communicability test, the subjects' task is an abbreviated version of the pencil-and-paper game "battleship" [5]. In the British free conversational test, subjects are given some task such as the comparison of pairs of photographs that includes the participants to talk for about 10 min.

POTENTIAL APPLICATIONS

The exact nature of digital data transmitted with the voice data would vary from one operating environment to another. In a tactical environment where relatively brief conversation is transmitted over a half-duplex link, digital data could carry only a few written words and/or simple hand-drawn graphics. On the other hand, in an office-to-office environment where relatively long conversations are transmitted over a full-duplex link (which is opened even during the listening period), more extensive digital data could be transmitted. The following is a list of information which could be transmitted by the integrated voice/data system to aid voice communications.

- *Hand-scribbles such as a map indicating the enemy location*—A picture could replace many spoken words ("a picture is worth a thousand words"). This is an ideal application for tactical communications.

- *Typed numbers such as enemy coordinates*—Numbers are easy to forget, particularly when heard in a high-tension environment. The talker can simultaneously transmit typed numbers during conversation. This is another useful tactical application.

- *Typed words or phrases*—When the listener has difficulty hearing because of excessive background noise in a tactical platform, transmission of critical words is particularly useful.

- *Typed memo*—A short prepared memo may be sent during continuous conversation over a full-duplex link from one office to another. When no one answers the call, the system could automatically leave a short message.

- *The speaker's ID number and/or signature*—The speaker can transmit his (or her) ID number from the keyboard, or signature from the graphic device for speaker verification purpose.

There are other applications of digital data under voice. For example, it can convey certain speaker-specific information to improve LPC processed speech. The 2400-b/s LPC synthesizes speech by using the same canned excitation signal for all speakers. As a result, the recognizability of familiar voices over the 2400-b/s LPC is rather poor; only 69%, according to tests conducted at NRL [6]. The use of a speaker-dependent excitation signal would improve the speaker recognition. This area will be investigated in the future. In this report, digital data will comprise only texts or graphics.

TECHNICAL APPROACH

One way of achieving an integrated voice/data system is to multiplex voice and data into a single bit-stream. In fact, this approach has been used in an experimental *wideband* integrated voice/data system for packet networks [7]. The multiplexing approach is well-suited for packetized communications because the multiplexed bit-stream does not need a fixed rate. We, however, did not choose this approach because the integrated voice/data system must be interoperable with other 2400-b/s LPCs. Hence our approach is to replace perceptually less significant speech data from the LPC bit-stream with digital data. We will choose speech data in such a manner that their absence in the bit-stream will not degrade synthesized speech. Similarly, the presence of extraneous digital data in the bit-stream will not produce undesirable audio effects, such as pops, clicks, flutters, or warbles.

Exploitation of Time-Variant Speech Information

We note that not all encoded speech data are equally significant from a perceptual stand point. In other words, the rate of real speech information is variable from one sound to the next. For example, unvoiced speech (consonants) requires fewer bits to encode than voiced speech (vowels); silence requires even fewer bits to encode than either voiced or unvoiced speech. Figure 3 illustrates voiced and unvoiced frames which are usually intertwined even in a single syllable word. Therefore, the rate of speech information varies within a single word.

Encoding of voiced speech data requires more bits because the voiced speech spectrum often has four to five resonant frequencies (Fig. 4) requiring at least eight to ten LPC coefficients to represent them. Furthermore, each coefficient must be quantized accurately to minimize audible flutters in the synthesized speech of steady vowels. Actually, the 41 bits allocated by the 2400-b/s LPC to encode voiced speech are barely adequate, and no bits can be deleted.

On the other hand, encoding unvoiced speech requires fewer bits than voiced speech because it does not have predominant resonant frequency (Fig. 4). In addition, the unvoiced speech spectra (i.e., LPC coefficients) need not be quantized as accurately because we are accustomed to hearing the same unvoiced speech spoken differently from one speaker to another. The 2400-b/s LPC classifies silent period (i.e., gaps between words) as unvoiced frames. The deletion of bits from these unvoiced frames produces a negligible effect on the LPC output. Thus we transmit digital data unvoiced frames. To do this we delete some of the perceptually less significant speech data.

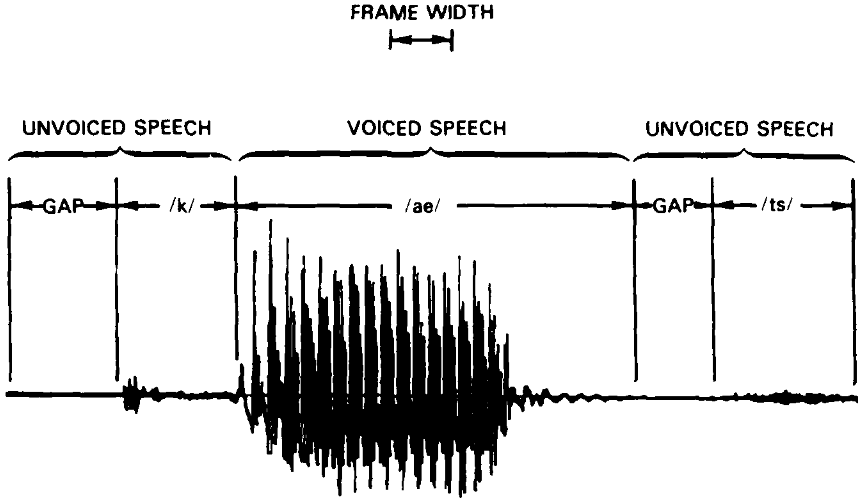
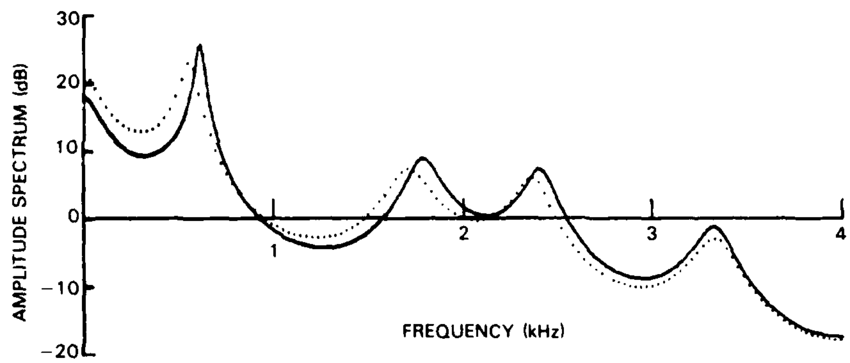
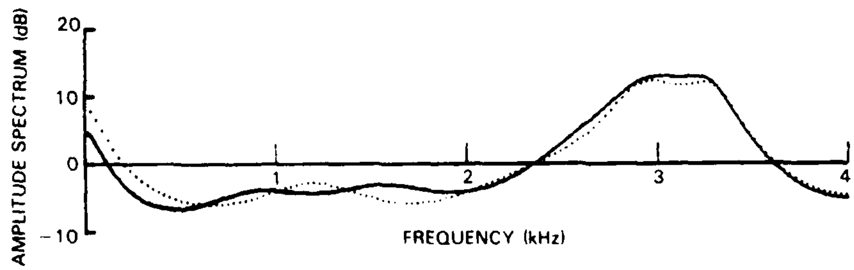


Fig. 3 — This trace shows the speech waveform of "cats" with the presence of voiced and unvoiced frames intertwined even in a single syllable word. The LPC analysis is performed once per frame (22.5 ms or 180 speech samples). The LPC classifies silence as unvoiced frames.



(a) Vowel /ae/ from "cats"



(b) Consonant /ts/ from "cats"

Fig. 4 — This shows the spectra obtained from portions of the speech waveform "cats" shown in Fig. 3. Solid lines are the spectra using quantized LPC coefficients defined by the 2400-b/s LPC

Unvoiced LPC Speech Data

The 2400-b/s LPC transmits the following 54 bits of data for each unvoiced frame [1].

- Four reflection coefficients describe the speech spectral envelope. Each coefficient is encoded by five bits, and the four most significant bits (MSBs) are protected by an (8,4) Hamming code. Thus the least-significant bit (LSB) of each reflection coefficient is not error protected.

- The root-mean-square value of input speech is transmitted for calibrating the amplitude of the synthesized speech. It is encoded by five bits, and its four MSBs are also error protected by an (8,4) Hamming code.

- The 60-pitch values and two-state voicing decision are combined into a seven-bit pitch/voicing parameter.

- The synchronization bit, as usual, is an alternating "0" and "1" from one frame to the next.

- There is one unused bit in each unvoiced frame.

Table 1 lists all the bits transmitted by the 2400-b/s LPC for each unvoiced frame.

Table 1 — Unvoiced Speech Data Transmitted by the 2400-b/s LPC

		SPEECH BITS					ERROR PROTECT BITS			
		LSB				MSB				
REFLECTION COEFFICIENTS	r(1)	0	1	2	3	4	5	6	7	8
	r(2)	0	1	2	3	4	5	6	7	8
	r(3)	0	1	2	3	4	5	6	7	8
	r(4)	0	1	2	3	4	5	6	7	8
RMS	A	0	1	2	3	4	5	6	7	8
PITCH & VOICING	P	0	1	2	3	4	5	6		
UNUSED BIT	D/C	0								
SYNC	SYNC	0								

Deleted Speech Bit

Not all bits listed in Table 1 can be deleted because they are sensitive to speech quality. These bits are discussed first.

- The synchronization bit does not contain speech information. It is essential for acquiring and maintaining frame synchronization. Thus it cannot be altered.

- The seven pitch/voicing bits behave as a group (for example, an unvoiced state is indicated by seven zeros). Hence none of the pitch/voicing bits will be changed.

- The 20 error-protection bits protect the four MSBs of reflection coefficients and the speech value. Thus none of the error-protection bits will be altered.

- The five bits will not be deleted because a good reproduction of unvoiced speech, particularly stop constants (i.e., /p/, /t/, /k/ and the like) are highly dependent on the accuracy of the speech value and its fluctuation from one frame to the next.

Some of the remaining 21 bits are not perceptually significant, and they may be reallocated for the transmission of digital data. For example, the unused bit in the bit-stream (Table 1) may be used freely because this bit is transparent to any of the 2400-b/s LPCs.

The remaining 20 bits represent reflection coefficients. Note that there is a wide range of speaker-to-speaker variations of unvoiced speech [8], and our brain is capable of discriminating unvoiced speech even if there is a substantial amount of spectral deviation. Thus the reflection coefficients from unvoiced frames need not be quantized accurately. Previously, we used only eight bits to represent unvoiced reflection coefficients in the implementation of a highly intelligible 800-b/s vocoder [9]. This indicates that there are some redundancies in the 20 bits allocated to encoded unvoiced reflection coefficients. Even if we delete the two LSBs from each coefficient, we still have 12 bits left. According to the diagnostic rhyme test (DRT), there is no appreciable speech degradation caused by the deletion of the two LSBs from each reflection coefficient (see the Experiment section of this report).

In summary, we have nine bits available from each of the unvoiced frames to transmit digital data. The deleted bits are indicated in the LPC unvoiced speech data (Table 2).

Table 2 — Unvoiced Speech Parameters Encoded by the 2400-b/s LPC

The bits indicated by shaded blocks are those which have been deleted to transmit digital data. Note that the bits indicated by hatched blocks are error-protected, whereas dotted block are not error-protected.

		SPEECH BITS LSB				MSB	ERROR PROTECT BITS			
REFLECTION COEFFICIENTS	r(1)	0	1	2	3	4	5	6	7	r
	r(2)	0	1	2	3	4	5	6	7	8
	r(3)	0	1	2	3	4	5	6	7	8
	r(4)	0	1	2	3	4	5	6	7	8
RMS	A	0	1	2	3	4	5	6	7	8
PITCH & VOICING	P	0	1	2	3	4	5	6		
UNUSED BIT	D/C	0								
SYNC	SYNC	0								

Digital Data Format

We group the digital data into 18 bits and transmit the digital data during two unvoiced frames at nine bits each. As indicated in Table 3, these nine bits are divided into four information bits and five status bits. The reason for choosing this data format is discussed next.

- *Information Bits*: Currently, computer data are structured in terms of 8-bit bytes. Thus it is natural to group eight information bits into one word. These eight information bits are transmitted over two unvoiced frames at four bits each. The two unvoiced frames must be such that the first frame has a sync bit of a logic "0," and the second frame has a sync bit of a logic "1." Thus certain unvoiced frames, frequently the one following voiced frames, will not carry digital data.

- *Status Bits*: The presence or absence of digital data in the unvoiced bit-stream is indicated by status bits. All status bits are set to "0" or "1" if digital data are absent or present, respectively. To transmit digital data under an error condition as much as 5%, we use as many as five identical status bits. The error probability of a status decision error is given in the next section.

Table 3 — One Word of Digital Data Consists of 18 Bits

STATUS BITS					INFORMATION BITS			
0	1	2	3	4	0	1	2	3

(a) When synchronization bit is "0"

STATUS BITS					INFORMATION BITS			
0	1	2	3	4	4	5	6	7

(b) When synchronization bit is "1"

Error Protection

All information bits are error protected by a Hamming (8,4) block code. The Hamming (8,4) code corrects all single-bit errors occurring within the eight-bit code word and also detects all double-bit errors.

We use five status bits to make the decision on the presence or absence of digital data. If we use a simple majority logic, the probability of making an erroneous status decision in terms of the independent individual bit-error of p, is

$$p_e = \binom{5}{3} p^3(1-p)^2 + \binom{5}{4} p^4 (1-p) + \binom{5}{5} p^5. \tag{1}$$

If the random-bit error of the channel is 5% (i.e., p=0.05), the status decision error probability, as obtained from (1) is 0.116% which is translated to one status decision error in 862 unvoiced frames. Since the frame rate is 44.44 Hz, and the number of unvoiced frames is approximately 40% of the total frame, one status decision error occurs once in 48 s on the average. That is an acceptably low status decision error rate

Integrated Voice/Data Bit-Stream

For voiced frames, the bit-streams of the 2400-b/s LPC and the integrated voice/data systems are identical. For unvoiced frames, the following changes will convert the 2400-b/s LPC bit-stream to the integrated voice/data system bit-stream:

- The speech bits occupying the least significant bits (LSBs) of the reflection coefficients and the unused bit (all indicated by dotted blocks in Table 2) are replaced by five identical status bits (all indicated by dotted blocks in Table 3). The status bits are "1"s if data are present, or "0"s if data are absent.

- The second LSBs of reflection coefficients (all indicated by hatch-lined boxes in Table 2) are replaced by the digital data (all indicated by hatch-lined boxes in Table 3): the four LSBs if the sync bit is a logic "0," or the four most significant bits (MSBs) if the sync bit is a logic "1." If the data are absent, these second LSBs will be speech data (i.e., no changes).

Expected Data Rate

The expected digital data rate is a product of the number bits per unvoiced frame (i.e., 4 bits) and the average rate of unvoiced frames. Certain unvoiced frames, however, frequently the one following voiced frames, will not carry digital data because two unvoiced frames in sequence do not necessarily have sync bits of "0" and "1," respectively. According to our observation with various speeches, about 3% of unvoiced frames are wasted.

As might be expected, the average rate of unvoiced frames (or selected unvoiced frames based on the above-mentioned criterion) varies from one speaker to another. It is significant to note that well-articulated speech with frequent gaps between words (as often heard from TV or radio broadcasters) has a higher percentage of unvoiced frames. On the other hand, slow speech (as often heard from Southerners) has actually a lower percentage of unvoiced frames because slowness of speech is very likely due to prolonged vowels, not due to prolonged gaps.

As indicated in Fig. 5, the number of *selected* unvoiced frames is somewhere between 40 to 50% of the total speech frames. Since we transmit 4 bits of digital data per unvoiced frame, and the frame rate is 44.44 Hz, the lowest data rate expected is $[(4)(.4 \times 44.44)] = 71$ b/s whereas the highest data rate expected is $[(4)(.5 \times 44.44)] = 88$ b/s. We estimate that we can transmit approximately 75 b/s during most speech.

Optional Fast Data Mode

So far, we have assumed that digital data can be transmitted at a rate of 4 bits per unvoiced frame *without* interrupting conversations. If we can interrupt conversations, we can transmit as many as 16 bits of digital data per unvoiced frame by deleting the four MSBs of each reflection coefficient. Each bit is still error protected, and the expected data rate is four times more, i.e., 300 b/s on the average. However, we did not incorporate this optional mode in our prototype device.

Input Data Buffer

Since we cannot transmit digital data during voiced frames, we must temporarily store incoming digital data. According to experimentation using a real-time device, an input buffer size of 256×8 bits (i.e., 256 words of digital data) is sufficient without experiencing an overflow of data.

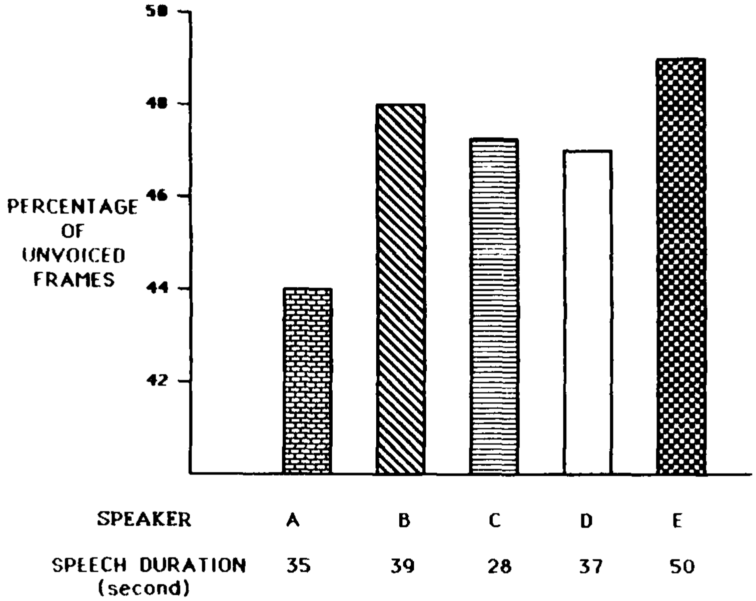


Fig. 5 — This shows a percentage of unvoiced frames in speech. To generate this graph, five different speakers casually read and recorded several paragraphs from newspapers (each read different articles), and recorded voices were then subjected to LPC analysis. The results show that nearly one-half of the speech frames are unvoiced (or silent). Note that the lowest percentage figure comes from a slow talker (i.e., Speaker A), and the highest percentage figure comes from a fast talker (i.e., Speaker E).

Text and Graphics Encoding

Among the eight information bits contained in one digital word (listed earlier in Table 3), one bit is used as an overhead data to indicate whether digital information represents characters or graphics (i.e., "0" if character, or "1" if graphics). The transmission of a character requires the remaining seven bits. As discussed earlier, the expected data rate is somewhere between 71 b/s and 88 b/s. Thus we can transmit approximately 10 characters during continuous conversation.

If the overhead bit is "1," the digital data contains graphics information. Any graphic device with a resolution of 512-by-512 points is more than adequate to transmit hand-scribbles. The initial point (i.e., the origin) will be encoded by 18 bits, 9 bits each for the x- and y-coordinates. Subsequent points are encoded differentially by 2 bits each. One bit is allocated to identify whether a given point is an origin or a differential point. The remaining 6 bits contain graphic information. Therefore, three digital words are needed to encode one initial point, whereas three differential points can be encoded by one digital word.

As an example, Fig. 6 shows the time required to transmit a simple hand-drawn figure. Figure 6 is a topographical sketch where an enemy concentration is indicated by "X." This figure contains four initial points and 303 differential points. Hence, the transmission of this figure requires $[(4 \times 3) + (303/3)] = 113$ digital words or $(113 \times 8) = 904$ bits. Because the average digital data rate is 75 b/s, it will take 12 s to transmit during continuous conversation.

EXPERIMENTATION

This section describes experiments related to the transmission of text or graphics with continuous speech. DRT scores of the LPC processed speech are also included.

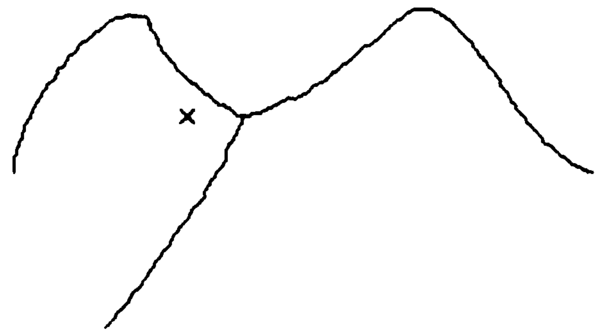


Fig. 6 — This is a hand-drawn chart with an enemy concentration indicated by "X." This could be a picture transmitted by the forward observer on a helicopter or aircraft who support the offshore Navy artillery offensive on suspected enemy ambush site near a mountain valley not far from a highway. With the use of the integrated voice/data systems, however, he can not only communicate verbally with the fleet, but also transmit this sort of visual aid. This figure, made of 307 points among which four are initial points, can be transmitted in 12 s without interrupting voice communication.

Prototype System

A real-time narrowband integrated voice/data system was devised by using the programmable 2400-b/s LPC resident at NRL. The programmable 2400-b/s LPC has a teletypewriter (TTY) attached by means of a RS232 port. The TTY is used for controlling terminal functions as well as debugging software. We used this RS232 port to transmit digital data.

Text Transmission

We transmitted both continuous speech (i.e., AM radio news) and continuous text generated by the computer at a rate of 10 characters per second or $(8 \times 10) = 80$ b/s through the integrated voice/data system. Although the average digital data rate in this experiment was 10 characters per second, the theoretical maximum rate during a speech gap is 22 characters per second since one character can be transmitted over two frames. (Note that the LPC frame rate is 44.44 Hz.)

To evaluate the quality of text under bit-error conditions, we introduced random bit errors of 0%, 0.5%, and 1% (Table 4). Since all information bits are error protected by an (8,4) Hamming code, the text has fewer errors than the channel errors. The text is highly legible at the error rates tested.

Compatibility with Other 2400-b/s LPCs

To verify the interoperability between the integrated voice/data system and the 2400-b/s LPC, we connected the voice/data system transmitter with a 2400-b/s LPC receiver. The processed speech sounded like that of a conventional LPC. Also, the same results were achieved when we connected a 2400-b/s LPC transmitter to a voice/data system receiver.

Speech Intelligibility

Despite eight bits of speech data being deleted from each unvoiced frame in the narrowband integrated voice/data system, apparently there is no loss of speech intelligibility as measured by the DRT (Table 5). The test is for three male speakers in a quiet environment.

Table 4 — Received Text at 80 b/s with Continuous AM Radio News

```
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPE OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED`OVER THE LAZY DOGS.
THE UUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QEICK BROW'I FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
```

(a) With 0% random bit error

```
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
```

(b) With 0.5% random bit errors

```
THE QUICK CROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK$BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVERTHE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUUMPD OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE DAZY DOGS.
THE QUICK BROWN FOX JUMPED OVR THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LA^Y DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
HE QUICK BROWN FOX JUMPED OVER HE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
THE QUICK BROWN FOX JUMPED OVER THE LAZY DOGS.
```

(c) With 1% random bit errors

Table 5 — DRT Scores of 2400 b/s LPC with and without 80-b/s Digital Data

Sound Class.	Without Data	With Data
Voicing	90.1	93.0
Nasality	94.3	97.7
Sustention	85.7	76.6
Sibilation	91.4	94.5
Graveness	80.5	78.9
Compactness	92.7	96.1
Average	89.1	89.5

CONCLUSIONS

All low-bit-rate voice encoders, including the 2400-b/s LPC, have been designed to transmit speech alone. This unfortunate design philosophy has been carried over from the telephone, which transmits only speech in analog form. As a result, there is no narrowband integrated voice/data system in existence.

This report shows that the simultaneous transmission of speech and digital data is feasible because: (a) the 2400-b/s LPC transmits speech information in the digitized form, and (b) the information rate of certain speech sounds are below the required transmission rate. Not only is this feasible, we can make the 2400-b/s LPC transmit both voice and digital data simultaneously *without* causing operational incompatibility with other 2400-b/s LPCs which do not have this feature.

Some useful applications of data under voice would be for the transmission of visual aids (written headlines, memos, or simple hand-sketched illustrations) to improve speech communications. We can transmit digital data at a rate of approximately 75 b/s without degrading speech intelligibility.

Voice communication plays a vital role in command and control of the armed forces. This added voice/data transmission capability could enhance the effectiveness of voice communications.

ACKNOWLEDGMENTS

The author thanks Robert Martin and Ralph Allen of Space and Warfare Systems Command who supported this task. The author also thanks Mark Lidd of Signal Processing Solutions, Inc. who generated the software for the real-time implementation using the programmable 2400-b/s LPC; this task could not have been completed without his assistance.

REFERENCES

1. "Analog to Digital Conversion of Voice by 2,400 Bits/Second Linear Predictive Coding," *Federal Standard 1015*, Published by General Services Administration, 7th and D Streets S.W., Washington, DC 20407, Nov. 28,1984.

2. G.S. Kang and S.S. Everett, "Improvement of the Narrowband Linear Predictive Coder"; Part 1—Analysis Improvements, NRL Report 8645, Dec. 1982; Part 2—Synthesis Improvements, NRL Report 8799, June 1984.

3. G.S. Kang and S.S. Everett, "Improvement of the Excitation Source in the Narrow-band Linear Prediction Vocoder," *IEEE Trans. Acoustics, Speech and Signal Proc.* **ASSP-33**, 377-386 (1985).

4. A. Schmidt-Nielsen and D.C. Coulter, "Effect of Modest Sidetone Delays in Modifying Talker Rates and Articulation in a Communications Task," Conference Record, pp. 493-96, Presented at the 97th Meeting of the Acoustic Society of America, MIT, Cambridge, MA, June 12-16, 1979.

5. A. Schmidt-Nielsen and S.S. Everett, "A Conversational Test for Comparing Voice Systems Using Working Two-Way Communication Links," *IEEE Trans. Acoustics, Speech and Signal Proc.* **ASSP-30**, 853-863 (1982).

6. A. Schmidt-Nielsen and K.R. Stern, "Identification of Known Voices as a Function of Familiarity and Narrow-band Coding," *J. Acoust. Soc. Am.* **77** (2), 658-663 (1985).

7. C.J. Weinstein, "Wideband Integrated Voice/Data Technology," Semiannual Technical Summary Report to the Defense Advanced Research Project Agency, Prepared by Lincoln Laboratory, Lexington, MA, Sept. 1984.

8. P. Steven, "Spectra of Fricative Noise in Human Speech," *Language and Speech* **3**, 202-219 (1960).

9. G.S. Kang and L.J. Fransen, "Low-Bit Rate Speech Encoders Based on Line-Spectrum Frequencies (LSFs)," NRL Report 8857, Jan. 1985.

END

FILMED

2-86

DTIC